DOCUMENT RESUME

ED 302 928 EA 020 610

AUTHOR Bain, Helen Pate; And Others
TITLE Three-Year Longitudinal Study of Small Class Size: The Metro-Nashville Public Schools Study: 1984-87.
PUB DATE Nov 88
NOTE 29p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (17th, Louisville, KY, November 8-11, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Attendance; *Class Size; *Cost Effectiveness; Longitudinal Studies; *Outcomes of Education; Primary Education; *Small Classes; Socioeconomic Status; *Student Behavior
IDENTIFIERS *Nashville Metropolitan Public Schools TN

ABSTRACT

Recent research suggests that small (1:15) classes in the primary grades improve a student's chance to obtain a solid educational foundation. Funding has proved a major hurdle to definitive class size research. Also, many class size studies are inconclusive, since few are longitudinal. Tennessee is weighing its burgeoning welfare and correctional costs against better educational outcomes possible through increased individual attention in smaller classes. In Fall 1984, Tennessee State University initiated a 3-year longitudinal study using an experimental (1:15) group versus a control (1:25) group. The study began in the first grades of two similar schools and followed cohorts of pupils moving through second and third grade. A "blind" control group was matched with the experimental group on five variables. Pupil test scores, self-concept, attendance, behavior, and teacher perceptions were analyzed using analysis of variance and co-variance techniques. In grade 1, important and statistically significant pupil achievement gains favored the experimental group, especially pupils of low socioeconomic status. By grade 2, achievement gains were less pronounced, and by grade 3 had evened out. There were no differences among groups in attendance. Although not statistically significant, self-concept continually improved in the experimental group, but showed no consistent pattern in the control or blind control groups. Behavior favored the experimental group in grade 1 and became significantly better in grades 2 and 3. In the first year, pupils got "up to speed" in smaller classes; over time, gains did not continually increase. The groups evened out by grade 3, but small-class pupils maintained what they had gained. Results clearly identified the need for a large-scale, longitudinal, class size study using the class, rather than the pupil, as the unit of analysis. Included are tables and 13 references. (MLH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED302928

THREE-YEAR LONGITUDINAL STUDY OF SMALL CLASS SIZE: THE METRO-NASHVILLE PUBLIC SCHOOLS STUDY: 1984-1987*

Helen Pate Bain, Associate Professor
Educational Administration and Supervision
College of Education
Tennessee State University
Nashville, TN 37203

C. M. Achilles, Professor and Chair
Educational Administration, Higher Education, and Research
Curry School of Education
University of North Carolina at Greensboro
Greensboro, NC 27412-5001

and

Mary Witherspoon-Parks
Metro-Nashville Public Schools
Bransford Avenue
Nashville, TN 37204

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

MSERA, 1988

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. P. Bain

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

*Supported in Part by the
Center for Educational Studies and Development,
University of North Carolina at Greensboro

EA020 610

BEST COPY AVAILABLE

Contents

Item	Page
Cover	i
Contents	ii
Abstract	iii
Introduction and Background	1
The Metro-Nashville Class-Size Study	3
Curriculum, Instruction and Teacher Factors	5
Strengths and Weaknesses: Limitations	5
Academic Achievement and Other Results	6
Academic Achievement	6
Self Concept	9
Attendance	10
Behavior	10
Classroom Activities	10
Further Analyses of Achievement	16
Summary of Findings	18
Study Follow-up: 1987-88	20
Conclusions	22
References	24

Tables

1. ANOVA (Pre and Post) and ANCOVA (Pre-to-Post) 1985	7
2. Summary of Comparison (ANCOVA) 1986	10
3. Summary of Comparison (ANCOVA) 1987	10
4. Comparisons of Punishment Proportions, Grades 2 and 3	12
5. Teacher Mean Scores on Classroom Conditions Inventory 1985-1987	14
6. Tests Between Free and Not Free Lunch Pupil Scores (Reading: Grade 1)	17
7. Tests Between Free and Not Free Lunch Pupil Scores (Math: Grade 1)	17
8. Summary Comparison of Reading/Math Comparisons (SES) Grades 2 and 3	19

Figures

1. Graph of Self-Concept Scores (X, O) for 3 Years	12
2. Items from Classroom Conditions Inventory	15

ABSTRACT

Three-Year Longitudinal Study of Small Class Size: The Metro-Nashville Public Schools Study

Among strategies to help students increase basic skills scores has been class-size reduction, an expensive alternative. Results of the many class-size studies are inconclusive, especially since few are longitudinal. In fall, 1984, Tennessee State University (TSU) initiated a three-year longitudinal class-size study (1:15 vs. 1:25).

"Experimental" (X) and "control" (O) schools served similar neighborhoods. A "blind" control group (O_b) was matched with (X) on 5 variables. Pupil (CAT, Stanford) test scores, self concept, attendance and behavior, and teacher perceptions were analyzed with ANOVA at pre and post and ANCOVA for pre-to-post.

In grade one, important and statistically significant pupil achievement gains favored (X), and especially low SES pupils. By grade 2, achievement gains were less pronounced; by grade 3, achievement evened out.

There were no differences among groups in attendance. Although not statistically significant, self concept continually improved in (X), but showed no consistent pattern in (O) and (O_b). Behavior favored (X) in grade one and became significantly better ($p \leq .05$) in grades two and three.

In the first year, pupils get "up to speed" in smaller classes, but over time gains do not continually increase. The groups evened out by grade three, and small-class pupils maintained what they gained.

Three-Year Longitudinal Study of Small Class Size: The Metro-Nashville Public Schools Study

Introduction and Background

As educators have sought ways to meet policy mandates for improved pupil outcomes and increased mastery of "basics," they have revisited and dusted off some previously considered ideas. One of these ideas has been class size. The decision to reduce class size (or teacher-pupil ratio) is important because it is costly and will establish precedents that will not be easy to change. The early 1980's saw a renewed interest in class size and a new generation of class-size studies.

Although class size has been debated for years and logic cries that reduced class size should benefit pupils and teachers, the debate is still much with us. One classic debate focused on a meta-analysis of class size studies (Glass, et al., 1979; Education Research Service, 1978, 1980). A review by Tomlinson (1988) suggests that reducing class size may not be the answer to improving education. Still, some research suggests that small (1:15) classes in the early grades improve a student's chance to get a sound educational foundation. This, in turn, greatly increases the possibility that a student will complete high school without failure. Early class-size studies debated different class sizes from those of concern today; they often studied reductions from 36 to 25 pupils or so. Balow (1969) found that a class-size reduction from an average of 1:30 to 1:15 seemed to lead to achievement gains in reading for pupils in grades one to four. Perhaps here is one of the problems. The real payoff, some suggest, will only come with a significant class size reduction--one teacher with up to 15 pupils.

Helen P. Bain, Tennessee State University, Nashville, TN 37203; C.M. Achilles, University of North Carolina at Greensboro, Greensboro, NC 27412-5001.
Portions of this material are adapted from Whittington (1985), and Whittington, Bain and Achilles (1985).

A major hurdle for definitive class-size research is funding. A classroom with one teacher and 15 pupils (1:15) is more expensive than 1:25 or so. Government and policy groups, grappling with rampant costs, may summarily reject smaller class sizes as a too costly use of public funds. This perception of excessive present cost fails to consider potential savings if small classes are highly successful over the long haul. Suppose, after being in small classes, students did not have to repeat grades later on. What would the savings be if there were fewer dropouts, less truancy, fewer discipline problems and reduced future welfare payments? In Tennessee in 1983-84, an incarcerated youth cost the state $21,170; an adult (in 1985) cost $15,027. Approximately 54,000 pupils were retained in Tennessee in grade (1984), requiring 2,160 teachers at the state's ratio of 1:25. The state's welfare program alone is $429 million for 581,000 persons. Someone should weigh these correction and welfare costs against better education outcomes to put the education expense issue in perspective. Just suppose that increased individual attention possible in 1:15 classes (an investment) could help reduce welfare and prison costs and attain such altruistic goals as better citizenship and skilled labor.

Renewed interest in small primary grade classes is spreading slowly. The new generation of studies is analyzing longitudinal issues in class size, something of concern to many critics of the prior "one-shot" studies. In 1981 Indiana began Prime Time, a two-year experiment with 24 K-2 classes of 1:14 in nine schools. Study results quite impressively favored the small classes. Based on the results of the pilot study, the Indiana legislature endorsed a program to reduce class size in K-3 to a 1:18 ratio. It began with first grade in 1984-85 and included second, third and kindergarten in succeeding years (Mueller, et al., 1987).

In the fall of 1985, supported by the legislature's mandate for a class-size study and support of $3 million per year for a four-year longitudinal

project, the Tennessee State Department of Education began Project STAR (Student Teacher Achievement Ratio). Some of the impetus and direction for the state-wide STAR Project came from the Metro-Nashville Public Schools Class-Size Study reported here (Whittington, et al., 1985).

The Metro-Nashville Class-Size Study

In September, 1984, Tennessee State University's (TSU) Center for Excellence in the Teaching of Basic Skills to Economically and Educationally Disadvantaged Students initiated the Class-Size Study in the Metro-Nashville School System. The project continued until June, 1987. The three-year longitudinal study began in the first grades of two similar schools (1984-85). The cohorts of pupils moved to grade two (1985-86) and grade three (1986-87) with different teachers each year. The project personnel analyzed pupil and teacher variables in the different classes.

Student variables include: (1) reading and math achievement; (2) behavior; (3) attendance; and (4) self-concept. Other purposes were to identify how class size might influence teaching and learning and to review teachers' perceptions of classroom conditions. Although not a primary purpose, staff sought to determine if reduced class size seemed to affect different groups differently (pupil achievement by race, sex, SES, etc.).

Metropolitan Nashville (Metro) Public School System pupils in grade one (1984-85) constitute the population. The "sample" was grade-one pupils in two schools serving similar neighborhoods. One school contained the experimental or 1:15 classes, and the second school served as a "control" (1:25). Schools had space for the project, and appropriate agreements to participate were completed, including parental consent (for pupils).

The experimental group (X) consisted of 105 pupils in seven classes (1:15) beginning in 1984 in first grade. The immediate school community consisted of predominately single-family residences and trailer parks.

The control group (O) was from a comparable neighboring elementary (K-6) school. Grade one (1984) included 90 pupils divided among three and one-half teachers with a class size of 25 students, or normal class size.

The blind control group (O_b) of first-grade students (n=105) was drawn from 35 elementary schools throughout Metro and was matched with the (X) group demographically according to the following five criteria: (1) sex; (2) race; (3) economic status (Federally subsidized lunch); (4) birthdate within 45 days; and (5) California Achievement Test, Level 10 (CAT 10) total pre-reading raw score within four points.

The study really began with two control groups and one comparison group. The primary control group (O) was a nearby school. The "blind control group" (O_b) was drawn from the rest of the approximately 5,000 Metro grade-one pupils. Comparable results (e.g., city-wide testing) of all Metro grade-one pupils served as a point of comparison. When Metro changed its testing program in 1986, this comparison group was less useful in years two and three.

The (X) and (O) groups were comparable within the Metro System according to the following characteristics: (a) middle to lower-middle class community; (b) racial mix acceptable to the system's court-established ratio of 80 percent white and 20 percent black (achieved by busing from inner-city housing projects); and (c) generally comparable student achievement over time. For example, in 1983-84, the two schools and Metro had the following average raw scores for pupils in grade two. The (O) and Metro scores exceeded the (X) school's scores only slightly on both reading and math.

	Experimental (X)		Control (O)		Metro	
1983-84	Math	Read.	Math	Read.	Math	Read.
CAT 12 (grade 2)	47.8	47.3	54.0	53.0	50.7	51.5

The (X) school was some below the (O) school on one variable. In the (X) school, 57 percent of students received federally subsidized lunches; the (O)

school had 39 percent, and Metro had 45 percent. The (X) school enrolled 23 percent black pupils, and the (O) school enrolled 26 percent black pupils.

Curriculum, Instruction and Teacher Factors

All groups followed the adopted Metro-Nashville elementary curriculum. Each individual teacher determined instructional methods. Teachers in the (X) school kept logs of impressions, perceptions, instructional activities and teaching methods employed during the school year. No instructional interventions were provided during this study, and no extra support, supplies, materials, etc. were to provided to any group. The teachers each year were those employed by Metro and assigned to the schools. The project worked with only those personnel assigned to the schools in the usual way. All held at least the Bachelor's degree and were appropriately certified. The setting was kept as natural as possible in order to isolate, as well as possible, the impact of reduced class size (1:15).

Strengths and Weaknesses: Limitations

The study operated as a field study, subject to the realities of conducting research in schools as they usually operate. Researchers had no control of teacher selection, training, etc., or of pupils. There was no randomness, and the project operated totally within Metro guidelines. Pupil mobility was a problem. Each year new students had to be added to replace those who left.

The blind control (O_b) was an attempt to "match" the (X) group and mitigate that there were too few classrooms to use the classroom as the unit of analysis. Only (X) teachers kept logs in year one; (X) and (O) kept logs in years two and three. The act of reflecting and writing may have influenced some outcomes; (O) teachers should have kept logs in the first year.

The longitudinal nature of the study overcame a major criticism of prior class-size studies. The lack of randomness, no true control group, and inability to use the classroom as the unit of analysis (too few classrooms) mean that

generalizations from these results must be made with care, in context, and as part of findings of the stream of class-size studies. Since there were better control, less mobility and stronger use of the (O_b), analyses for grade one were more straightforward and more detailed than analyses in years two and three.

Academic Achievement and Other Results

Academic Achievement

The California Achievement Tests (CAT) were used in grade one because Metro was using (in 1984-85) CAT system-wide. The CAT allowed comparison of project math and reading scores with all Metro first-grade students; more importantly, this choice allowed the use of a blind control group. This plan was confounded when Metro changed its district-wide testing program in 1986 from CAT to Stanford. Researchers continued to give (X) and (O) pupils CAT in second grade for continuity, but (O_b) and system comparisons were disrupted.

To analyze first-grade student achievement data, researchers chose a combination of one-way analysis of variance (ANOVA) comparing pretest scores for three groups, (X), (O) and (O_b) and also an ANOVA on posttest scores for the same groups. Analysis of covariance (ANCOVA) using the pretest (CAT 10) as the covariate was used to analyze pre-to-post gains. Table 1 shows, for the three groups (X, O, O_b), results of the ANOVA for pretests and posttests, and the ANCOVA (pre-to-post) for grade one. Tests of homogeneity of variance showed that variances were similar and acceptable for all ANCOVAs in this study.

At pretest, the (X) group was statistically different from (O), but not (O_b) in Reading. The (O) and (O_b) groups were also different ($p \leq .05$, Sheffe test) on Reading. The pretest ANOVA (Math) showed that the (X) group was not statistically (.05 Sheffe) different from the (O) and (O_b) groups, but that (O_b) was statistically ahead of the (O) group. At posttest (ANOVA, Reading), the (X) group was significantly ahead of both (O) and (O_b) groups. The posttest ANOVA (Math) showed that (X) was significantly ahead of (O), but not (O_b). There were

TABLE 1

ANOVA (PRE AND POST) AND ANCOVA (PRE TO POST) RESULTS FOR THREE GROUPS (X, O, O_b) FOR 1:15 STUDY ON READING AND MATH SCORES

	Reading Test Results								Math Test Results							
	ANOVA						ANCOVA		ANOVA						ANCOVA	
	CAT 10 (Pre)*			CAT 11 (Post)			Pre to Post		CAT 10 (Pre)			CAT 11 (Post)			Pre to Post	
Group	N	$\bar{x}$	Rk**	N	$\bar{x}$	Rk	N	$\bar{x}$	N	$\bar{x}$	Rk	N	$\bar{x}$	Rk	N	$\bar{x}$
1. (X)	80	102.90	1	72	46.9	1	72	46.9	79	18.0	2	72	46.9	1	71	47.1
2. (O)	66	9: 24	3	82	38.5	3	65	39.2	65	17.4	3	81	42.3	3	64	42.3
3. (O_b)	80	102.86	2	66	38.9	2	66	38.9	80	19.5	1	66	44.0	2	66	44.0
Results																
F.		7.180			15.92			14.14		3.890			7.08			8.48
Sig.		.000			.000			.000		.022			.001			.000
Diff. at .05		3 vs. 2*** 1 vs. 2			1 vs. 2 1 vs. 3			N/A		3 vs. 2			1 vs. 2			N/A

*Groups X and O_b were matched on CAT 10 (Criterion was within four points). Note that reading CAT 10 had 120 items and CAT 11 and 60 items, providing lower mean scores on CAT 11. CAT 10 (math) had 26 and CAT 11 had 56 items.

**RK=Rank ***Sheffe Test for difference between pairs; refer to groups above.

no significant posttest differences between the two control groups (O, O_b) on either Reading or Math. This finding encouraged the use of ANCOVA.

ANCOVA analyses showed that pre-to-post gains were considerably different among groups ($p \leq .000$) for both Math and Reading. The mean scores show that results clearly favor (X); both control groups (O, O_b) were similar and lower than the (X) group. (There are fewer pupils for the ANCOVA analysis than for ANOVA since only those with all four test results were used.)

An additional, non-statistical indicator favoring (X) is the ranking of mean scores at both pre and post. The (X) group, although not highest at the beginning, was highest at the end. Posttest analyses favored (X), and the ANCOVA indicated that the pre-post gains were statistically important.

Based on the ANCOVA and ANOVA analyses, the (X) group consistently achieved better than either (O) or (O_b). Since the only treatment was class size reduction (excluding such factors as the halo effect), one inference is that reducing class size to 1:15 in grade one had a positive effect on pupil reading and math outcomes in this study.

In second grade (year two of the study), the Metro system changed to the Stanford test series. In grades two and three, the researchers were able to use analysis of covariance (ANCOVA) with the Stanford results (Primary II and III for second and third grades respectively) as the outcome and the CAT 10 as the covariate. This meant that this particular analysis could include only those pupils who began the project in 1984 and had CAT 10 scores and also had the required outcome scores. The numbers of pupils upon whom these analyses could be conducted would reduce considerably. In second grade those numbers were: X=61; O=32 and O_b=39. In third grade the numbers were: X=40; O=24 and O_b=40.

As Table 2 shows, there were no differences among groups (X, O, O_b) on grade two tests--all three groups had essentially the same group mean scores and relative gains from CAT 10. Table 3 shows that (O), with a group mean of 68.6, had, by grade 3, moved ahead of (X) and (O_b) with means of 63.7 and 66.6 respectively in reading, but that there were no differences in the math achievement gains (pre to post; ANCOVA).

These three-year results show that there was considerable gain in grade one favoring (X) and that this gain seemed to be maintained throughout grades two and three [except the reading ANCOVA with the (X) and (O) groups in grade three]; the (X) group did not make any additional gains over those achieved in the first year. Pupil cognitive gains, however, were not only the only indicators of project success.

Self-Concept

A modified Piers-Harris Children's Self-Concept Scale was used to assess self-concept. Although originally designed for students in grades 3-12, the scale was adapted for use with younger children. Adaptation included reducing the items from 80 to 34, simplifying vocabulary, and administering the scale orally and individually. Adaptation reliability was estimated by retesting one class three weeks after the first testing [Test-Retest = .70 (Sig. $p \leq .003$); Cronbach alpha (pre) = .83; (post) = .70]. Validity was checked by obtaining each teacher's perceptions of the high and low self-esteem exhibited by students in her class.

The pre and post self-concept testing indicated no significant difference between the (X) and (O) groups in grade one. Children were tested orally and individually by the same tester. A school-wide self-concept program already in place at the (O) school could have influenced results, but this is only conjecture. Testing of self-concept at this age can be affected by the children's immaturity. For grades two and three, self-concept was tested only in the

TABLE 2

Summary Comparison of Stanford Primary II Total Math and Total Reading Achievement: ANCOVA, Metro Class-Size Project, 1986

	Groups				
	X	O	O_b	F	P
n =	60	30	39		
Read. x =	72.7	73.4	72.1	.11	.90
Math x =	74.5	75.7	74.1	.14	.87

Dependent Var. = Stanford Primary II Total Math (or Reading) Percentage Correct.
Covariate = CAT 10 Total Math (Reading) Correct.

Data from Dennis (1986, pp. 103 and 107).

TABLE 3

Summary Comparison of Stanford Primary III Total Math and Total Reading Achievement: ANCOVA, Metro Class-Size Project, 1987

	Groups				
	X	O	O_b	F	P
n =	40	24	40	3.36	0.04
Read. x =	63.7	68.6	66.6	.02	.98
Math x =	58.9	56.6	63.0		

Dependent Var. = Stanford Primary III Total Math (or Reading) Percentage Correct.
Covariate = CAT 10 Total Math (Reading) Correct.

Data from Parks (1988, p. 76).

spring of each year. In general, the average self-concept of the (X) group showed a slow but continuous rise, and the (O) group fluctuated considerably. At pretest (Fall, 1984), the (X) group score (24.8) was below the (O) group (26.3). At project end (Spring, 1987), the (X) group's average score (27.0) was higher than the (O) group's (24.7). These relationships are in Figure 1.

Attendance (All Years)

Attendance patterns were tabulated and compared through a review of seven monthly attendance reports for three groups, and for all Metro first-grade students. These tabulations were made each project year, and tests of proportions conducted. There essentially were no differences among the groups, and the patterns of attendance showed no fluctuations over the project's three-year duration. The 1:15 project did not change any extant attendance patterns.

Behavior (All Years)

Corporal punishment incidents were recorded. With only one reported corporal punishment incident in grade one, the behavior record (.01) of (X) was superior to records for (O) (.29), (O_b) (.13) and Metro (.16). This measure, however, is too sparse for serious comparisons. For grades two and three, there were enough incidents of corporal punishment for meaningful comparisons. Using a test of proportion, a Z score showed that there were significantly fewer cases (proportionately) of corporal punishment in the (X) than in (O) or Metro (total). These results are in Table 4.

Classroom Activities

Teachers at the (X) and (O) schools completed a 24-item, researcher-constructed Inventory of Classroom Conditions to record their perceptions of: (1) classroom climate; (2) non-instructional duties and activities; (3) classroom management/student behavior; (4) preparing and implementing instruction; and (5) teacher and student interaction/individualization. Classroom activities

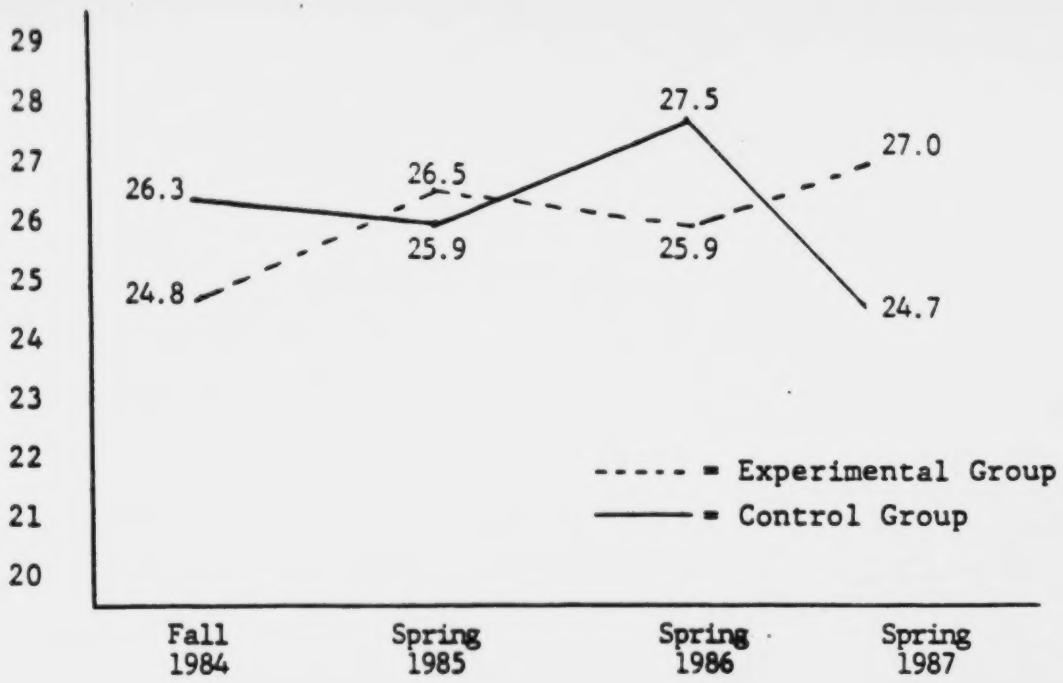

Figure 1. Graph of Average Self-Concept Scores of Regular (Control) and Small-Class (Experimental) Subjects, Fall 1984 to Spring 1987, in the Metro Class-Size Project

TABLE 4

Comparisons (Grades Two and Three) of Corporal Punishment Proportions: (X), (O) and Total Metro, Class Size Study, 1985-1987

Year/Group	Pupils	Incidents	Proportion	Proportion Test: Each Group vs. (X) Group Z	P
Grade Two					
X	85	9	.11	N/A	N/A
O	88	28	.32	-3.35	<.01
Metro	4990	878	.18	-1.67	<.05
Grade Three					
X	75	12	.16	N/A	N/A
O	89	96	1.08	-12.38	<.01
Metro	5083	1808	.36	-3.59	<.01

Sources: Grade Two, Dennis, (1986:143) and Grade Three, Parks, (1988:103).

were recorded in logs throughout the 1984-85 school year by (X) teachers, and in 1985-86 and 1986-87 by (X) and (O) teachers. Log entries were summarized in the same five categories of Inventory responses.

In all three years, the (X) (n=7) teachers scored considerably higher (translated as "better" or more positive) on 22 of 24 Inventory items, and markedly higher on key items relating to instruction and attention to the individual student (such as items 10, 11, 17) than did (O) teachers (n=4). These items reflected time to work with each student on individual interest and achievement, and opportunity to prescribe remedies for individual learning problems, to provide teacher-made materials, and to cover adequately all required material. Control-group teachers scored high and close to (X) teachers on items 21 and 23 (items of what might be done if they had fewer pupils in class). In 1984-85 and 1986-87, the (O) teachers scored higher than (X) teachers on these two items. Table 5 compares Inventory of Classroom Conditions average scores between (X) and (O) teachers on each Inventory item. (Data from Whittington, 1985; Dennis, 1986; Parks, 1988). Figure 2 shows the items on the Classroom Conditions Inventory.

The teacher log entries (grade one) indicated that reductions in class size can have a positive impact on both teaching processes and student behavior. Teachers believed that they were more productive, that children learned more at a faster rate, and that student behavior was better. The teachers agreed on the following points concerning the 1:15 situation compared to regular classes:

1. Classrooms were quieter with fewer student interruptions.
2. Students showed appreciation for each other, and all wanted to participate.
3. Potential discipline problems were identified and resolved quickly.
4. There were more learning activities with an increased amount of pupil participation.

TABLE 5

Mean Scores of Small-Class Teachers (n=7) (X) and Regular-Class (O) (n=4) Teachers by Years to the Classroom Conditions Inventory, Metro-Nashville Class-Size Study

	1984-85			1985-86			1986-87		
Item	X	O	Dif.*	X	O	Dif.*	X	O	Dif.*
1	7.2	6.8	.4	6.8	5.2	1.6	6.3	4.0	2.3
2	7.3	6.4	.9	6.8	3.2	3.6	7.3	4.0	3.3
3	7.0	5.5	1.5	6.4	3.1	3.3	7.0	2.0	5.0
4	7.0	5.1	1.9	6.4	2.5	3.9	6.5	2.0	4.5
5	7.2	6.5	.7	7.2	5.3	1.9	6.8	5.0	1.8
6	7.8	6.0	1.8	7.4	3.3	3.1	7.5	2.0	5.5
7	7.3	3.8	3.5	6.6	2.9	3.7	7.0	4.0	3.0
8	7.0	3.3	3.7	6.4	1.9	4.5	4.5	1.0	3.5
9	7.7	6.5	1.2	7.2	5.3	1.9	7.3	1.0	6.3
10	7.7	5.5	2.2	7.3	3.5	3.8	7.3	1.0	6.3
11	7.2	4.5	2.7	6.7	2.8	3.9	6.0	1.0	5.0
12	7.3	3.1	4.2	6.2	3.1	3.1	6.3	4.0	2.3
13	7.4	4.9	2.5	6.9	3.3	3.6	7.0	4.0	3.0
14	7.5	4.5	3.0	6.3	3.3	3.0	5.8	3.0	2.8
15	6.7	3.8	2.9	6.9	2.1	4.8	4.8	3.0	1.8
16	7.0	4.1	2.9	6.6	2.4	4.2	4.8	2.0	2.8
17	7.3	4.8	2.5	7.2	4.1	3.1	7.0	3.0	4.0
18	7.3	5.8	1.4	6.3	3.1	3.2	5.3	4.0	1.3
19	7.2	4.8	2.4	6.7	2.5	4.2	6.3	3.0	3.3
20	7.6	6.8	.8	7.1	1.9	5.2	7.0	5.0	2.0
21	7.8	8.0	-.2	7.2	7.0	.2	6.8	8.0	-1.2
22	7.5	3.5	4.0	6.8	2.2	4.6	6.3	1.5	4.8
23	7.3	8.0	-.7	7.2	7.0	.2	6.0	8.0	-2.0
24	7.7	4.1	3.6	7.3	3.4	3.9	7.5	5.0	2.5

*Dif.=Difference of Scores (X minus O)

Data from Whittington (1985), Dennis (1986) and Parks (1988)

Note: (X) exceeds (O) on all items (a more positive response favoring small-class size). The exceptions are items 21 and 23 where a lower (X) score or a smaller X-O difference would actually favor the (X) condition. Note also that differences (X minus O) generally are greater in 1985-86 and 1986-87 than in 1984-85.

1. My classroom is free of tension.

2. There is opportunity to interact with each student more than once during a lesson.

3. I have the opportunity to become familiar with home background, health history, and academic history of all my students.

4. There is sufficient time to diagnose individual student problems.

6. Routine tasks, e.g., calling roll, collecting money, require little time.

7. There is opportunity to work with each student on his/her individual reading interest and achievement.

8. There is time to provide teacher made materials for each student.

9. It is possible to rearrange my room to accommodate individual and small group activities and special projects.

10. Transition of activities, e.g., going to lunch, restroom, playground, takes little time.

11. When individual students interrupt the learning process, I have the opportunity for individual counseling.

12. There is opportunity to work with each student on his/her individual math interest and achievement.

13. It is possible to respond immediately to student emergency, physical-intellectual and behavioral needs (i.e., blow nose, go to bathroom, answer questions).

14. There is opportunity for me to prescribe remedies for individual learning problems (i.e., cutting skills).

15. It is possible to cover adequately all of the required material.

16. There is time to individualize adequately material for small groups.

17. It is possible to keep disruptions to a minimum in my class.

18. There is sufficient time to use a broad range of teaching strategies with my students.

19. I have the opportunity to explore the interests, attitudes, and self-development needs of my students.

20. I have the opportunity to deal with student disruptions in an appropriate way.

21. My classroom problems could be greatly diminished if class size could be 1 to 15.

22. I have enough time to listen to the problems and concerns of all of my students.

23. Mastery of basic skills could be greatly increased with a class size of 15.

24. I have enough time to keep discipline problems from erupting in my classroom.

Figure 2. Item List of Inventory of Classroom Conditions

5. There was more time to monitor students' on-task behavior during the instructional process; provide quicker and more thorough feedback to students; reteach concepts as needed; provide in-depth instruction.

6. Individualizing instruction significantly reduced the need for reteaching.

7. Individual student-student interaction helped students understand one another while increasing their desire to help each other.

Further Analyses of Achievement

Did small classes (1:15) have differential effects on various sub-grouping of pupils, such as low socio-economic status (SES), as noted by free lunch (FL) or not free lunch (NFL)? The researchers divided pupils (X, O, O_b) into the desired groups (FL, NFL) for further detailed analyses of the standardized measures of pupil achievement (CAT, Stanford).

At first grade, the statistical tests included t-tests of pretest and then of posttest reading and math scores within the (X), (O), and (O_b) groups for FL and NFL pupils and then the within-group ANCOVA to show pre-to-post gains on reading and on math for the FL and NFL pupils. Tables 6 (reading) and 7 (math) show that the slight pre-test differences favoring NFL in all three groups (X, O, O_b) were not statistically significant. By posttest, however, the differences favoring NFL were statistically significant in (O) and (O_b) comparisons (except O_b math which was .095). This supports the "stereotype" that higher SES pupils do better than lower SES pupils on cognitive standardized tests in schools. Note, however, that in the small-class (X) the less than one-point difference between FL and NFL at posttest does not approach significance. The ANCOVA analyses support the t-tests by showing that, except for (X) where there were essentially no differences ($p \geq .90$), the NFL group outperformed the FL group pre-to-post ($p \leq .05$) in reading and math.

Small class size (1:15) in grade one seems to help low SES pupils overcome the stereotypic lack of school readiness usually associated with low SES. The

TABLE 6

1:15 Project t Tests Between Free and Not Free Lunch Pupils on Reading Pre and Post Tests (1985) and ANCOVA (Pre to Post)

Reading		t-Tests									ANCOVA			
		Pre					Post							
Group	N n	$\bar{x}$	SD	t	Sig	N n	$\bar{x}$	SD	t	Sig	N n	$\bar{x}$	F	Sig
X	80					72					72	46.90		
1	34	100.88	17.43			29	46.31	8.08			29	46.31		
				-1.03	.31				-0.49	.62			0.27	.96
2	46	104.39	11.00			43	47.30	8.81			43	47.30		
O	64					78					63	39.17		
1	35	92.91	22.43			43	36.26	10.59			34	36.32		
				-0.36	.72				-1.92	.06			5.76	.019
2	29	94.97	22.89			35	41.17	11.71			29	42.52		
O_b	80					66					66	38.94		
1	34	100.88	17.57			27	35.11	9.37			27	35.11		
				-1.00	.32				-2.68	.01			6.95	.011
2	46	104.33	11.35			39	41.59	10.03			39	41.59		
TOT	224					216					201	41.87		
1	103	98.17	19.50			99	38.89	10.64			90	38.18		
				-1.66	.10				-3.24	.001			8.16	.005
2	121	102.11	15.23			117	43.56	10.47			111	44.05		

1=Free Lunch; 2=Not Free Lunch. CAT 10=120 items; CAT11=60 items. X=Exper.; O=Control; O_b=Blind Control.

TABLE 7

1:15 Project t Tests Between Free and Not Free Lunch Pupils on Math Pre and Post Tests (1985) and ANCOVA (Pre to Post).

Math		t-Tests									ANCOVA			
		Pre					Post							
Group	N n	$\bar{x}$	SD	t	Sig.	N n	$\bar{x}$	SD	t	Sig	N n	$\bar{x}$	F	Sig
X	79					72					71	47.06		
1	34	17.91	3.87			29	46.83	7.19			29	46.83		
				-0.18	.86				-0.12	.90			.11	.92
2	45	18.07	3.85			43	47.02	5.97			42	47.21		
O	63					77					62	42.23		
1	34	16.82	4.41			42	40.19	7.89			33	39.64		
				-0.72	.48				-2.22	.03			6.487	.013
2	29	18.03	8.12			35	44.46	8.80			29	45.17		
O_b	80					66					66	43.98		
1	34	19.35	4.92			27	42.04	8.07			27	42.04		
				-0.32	.75				-1.70	.095			4.73	.03
2	46	19.67	3.65			39	45.33	7.28			39	45.33		
TOT	222					215					199	44.53		
1	102	18.03	4.50			98	42.66	8.16			89	42.71		
				-1.00	.32				-2.83	.005			8.922	.003
2	120	18.68	5.17			117	45.69	7.36			110	46.01		

1=Free Lunch; 2=Not Free Lunch. CAT 10=26 items; CAT 11=56 items. X=Exper., O=Control, O_b=Blind Control

FL youngsters in (X) gained relatively more than FL youngsters in (O) and (O_b), and the (X) group outperformed both (O) and (O_b).

Analyses for FL/NFL differences in grades 2 and 3 used only those pupils who had the requisite CAT-10 scores and corresponding Stanford Primary II (grade 2) or Stanford Primary III (grade 3). Mobility and absenteeism reduced the (n) of pupils in the comparison groups considerably. An Analysis of Covariance (ANCOVA) was employed, using the CAT-10 as covariate. Summary results are in Table 8 for grades 2 and 3.

Although in grades 2 and 3 the NFL group consistently scored some higher than FL in the (X) and (O) groups, these differences never reached statistical significance. In all four comparisons between FL and NFL in (O_b), the NFL outperformed the FL, and in two of the four (grade 2 reading and math), the differences were statistically significant.

The SES comparisons (FL/NFL) support the regular achievement analyses in that there were positive gains for (X) in grade one [and also for FL youngsters in (X) in grade one], but that in grades 2 and 3 these differences do not continue. The next section contains information about possible explanations for study results and some follow-up steps taken by the researchers in 1987-88 to try to understand the results.

Summary of Findings

In grade one, important and statistically significant pupil achievement gains favored (X), and especially low SES pupils. By grade 2, achievement gains were less pronounced; by grade 3, achievement evened out.

There were no differences among groups in attendance. Although no statistically significant, self concept continually improved in (X), but showed no consistent pattern in (O) and (O_b). Behavior favored (X) in grade one and became significantly better ($p \leq .05$) in grades two and three.

TABLE 8

Summary Comparison of Reading and Math Achievement of Free Lunch (FL) and Not Free Lunch (NFL) Pupils (ANCOVA, Pre-to-Post) for Grades 2 (Stanford Primary II) and 3 (Stanford Primary III) with CAT-10 as Covariate, Metro-Nashville Class-Size Study (1984-1987)

		Math Grade 2 (Dennis, p. 111)				Reading Grade 2 (Dennis, p. 109)			
		FL	NFL	F	P	FL	NFL	F	P
X	n	22	38	.44	.51	23	38	.27	.61
	x	71.6	73.7			72.5	70.5		
O	n	13	15	.97	.33	14	16	.75	.39
	x	73.0	76.7			70.0	73.9		
O_b	n	14	25	8.94	.01	14	25	4.38	.04*
	x	69.5	82.0			68.6	77.6		

		Math Grade 3 (Parks, p. 84)				Reading Grade 3 (Parks, p. 80)			
X	n	15	25	.52	.48	15	25	.15	.70
	x	55.6	60.8			59.7	66.0		
O	n	11	12	2.27	.15	11	12	.56	.47
	x	54.0	61.6			67.1	71.3		
O_b	n	15	26	.02	.88	15	25	.05	.83
	x	60.4	64.5			63.7	68.3		

Data from Dennis (1987, pp. 109 and 111) and Parks (1988, pp. 80 and 84).
X=Experimental (1:15); O=Control (1:25); O_b=Blind Control.
Only pupils with all scores (pre-post) for each subject are included in the analyses.

*Significant at $p \leq .05$.

In the first year, pupils get "up to speed" in smaller classes, but over time gains do not continually increase. The groups evened out by grade three, but small-class pupils maintain what they gained.

Study Follow-Up: 1987-88

Class size remains a live issue (e.g., Tomlinson, 1988). Class-size reduction as a policy issue is expensive, and class-size studies have continued to be inconclusive. Thus, since the researchers were close to this study for four years: 1) they did some follow-up with teachers and pupils; and 2) they have speculated about the results obtained in this study. This section provides a summary of efforts to explain and understand more about the outcomes of the study. This section is speculative and goes beyond the actual findings.

It is possible that the major effects of small-class benefits to pupil gain in achievement occur in the first year. Perhaps the pupils generally "get up to speed," and it may be unrealistic to expect that they will continue to outperform other groups. (There is some support for this idea in some remediation efforts, especially the Reading Recovery Project in Ohio.) If this idea is true, it argues for small classes in first grade as a strategy to help pupils early on with "the basics."

In this (Metro-Nashville) study, the (X) pupils were all in a relatively low SES school (57 percent free lunch to a system-wide factor of 45 percent and a control school of 39 percent). The O_b pupils were from 35 separate schools throughout the total system, and some were from high SES situations. Do the concentration of low SES youngsters in one class (X) and the dispersion of the (O_b) youngsters in many classes influence the relative O_b gains in grades two and three? Follow-up interviews in 1988 with teachers of the top 10 and bottom 10 pupils in both math and reading (seven students accounted for 14 of the possible 20 top scores by having high scores in both reading and math) noted

that, in general, many (X) pupils were "at risk," with unstable family situations (divorce, working single parents), but that many (O_b) pupils came from stable homes and families that were involved in the schools. These data (7/88) are still being analyzed.

There was some control of pupil attendance at tests in (X) and (O), but in (O_b) there was none (it was a blind control group). There is evidence that some of the poorer academic pupils from (O_b) did not take the tests, creating a selective bias in favor of the better (O_b) pupils. Specifically, in grade two, only 24 of 39 (O_b) pupils took the Stanford test. Nine of the (X) pupils whose (O_b) matches did not take the test achieved raw scores from 91 to 113 [the average (X) raw score was 85]. These nine high scores were excluded from the matched-pair analysis. (This is still being analyzed.)

A limitation of this study, use of pupils rather than classes as the unit of analysis, may influence the results. A longitudinal class-size study with enough separate classes of (X) and (O) is expensive and complicated. If teachers make a difference, and teacher-pupil and pupil-pupil interactions influence learning, then the classroom, not individual pupils, must be the unit of analysis in improved class-size studies. Tennessee is currently supporting a longitudinal class-size study of sufficient magnitude to have the class as the unit of analysis. Project STAR, for grades K-3, is currently in its fourth year and working with the cohort of pupils in grade three. Tentative results show that there are substantial teacher effects.

The environment for teaching may be important. In the (X) condition, pupils in grades two and three were taught in an "open" physical space. Teachers and pupils (X) complained some of distractions and noise in grades two and three. The (O) school environment was traditional classes, as were most (O_b) environments.

The impressive gains of (X) in grade one are noteworthy. The researchers continue to wonder if the average scores of (X) in grades two and three may have been slightly depressed because the 1:15 condition allowed some students to pass who might, in the 1:25 condition, normally have been retained in grade one. If these marginal students had been retained and removed from the study, the (X) average would have been higher. Supporting evidence for this conjecture comes from the (O_b) group where, due to attrition, new pupils had to be identified for years two and three.

In 1986 Metro changed the month for administering the Stanford Primary III from April to January. This change deprived the treatment group (X) of three months of 1:15 instruction prior to the exit test.

Finally, there was a change of administration in (X) school just as the second year of the study began. A teacher served as acting principal for two months before the replacement was assigned. This disruption caused a period of adjustment for (X) faculty and students.

These items of discussion are not meant to explain away any specific findings. Researchers close to a study have a responsibility to raise additional questions and to go well beyond the mere reporting of numbers. Studies in schools must accept limitations on the desired "experimental" design. Thus, loose ends are normal in these types of studies, and researchers, it seems, should at least speculate on the questions left unanswered.

Conclusions

The Metro-Nashville class-size study provided some interesting findings, especially as the study followed a cohort of pupils for three years. As with some prior class-size studies, the results were less than conclusive, but the longitudinal nature did raise some new insights and questions. The favorable effects for (X), and especially low SES pupils, were clear in grade one. Over

three years the behavior of (X) pupils was better than that of (O) and the system-wide norm. Self-concept of (X) moved consistently upward. Teachers favored small classes, and their log entries defined more orderly, instructionally-oriented classes where teachers had time to teach and work with individual pupils. Results of this small study clearly identified the need for a large-scale, longitudinal, class-size study using the class (rather than the pupil) as the unit of analysis.

Bibliography

Achilles, C.M., H.P. Bain, F. Bellott, J. Folger, J. Johnston, M.N. Lintz. (1988). Student Teacher Achievement Ratio (Project STAR): Technical Report, Year One. Nashville, TN: Tennessee State Department of Education. March.

Balow, I.H. (1969). "A Longitudinal Evaluation of Reading Achievement in Small Classes." Elementary English. February, pp. 186-87.

Bain, H.P. and C.M. Achilles. (1986). "Interesting Developments on Class Size." Phi Delta Kappan, 67, 9, May, pp. 662-665.

Dennis, Ben. (1986). "Effects of Small Class Size (1:15) on the Teaching/Learning Process in Grade Two." Unpublished Doctoral Dissertation. Tennessee State University, Nashville, TN. August.

Education Research Service. (1978). Class Size: A Summary of Research. Arlington, VA: Education Research Service (ERS) Inc.

Education Research Service. (1980). Class Size Research: A Critique of Recent Meta-Analysis. Arlington, VA: Education Research Service (ERS) Inc.

Glass, Gene V., L.S. Cahen, M.L. Smith, N. Filby. (1979). "Class Size and Learning: A New Interpretation of the Research Literature." Today's Education. April/May.

Glass, Gene V., et. al. (1982). School Class Size: Research and Policy. Beverly Hills, CA: SAGE Publications.

Mueller, D.J., et. al. (1987). Project Prime Time. Final Report of Indiana's Class Size Experiment. Bloomington, IN: Indiana University.

Parks, Mary W. (1988). "Effects of Small Class Size (1:15) on the Teaching/Learning Process in Grade Two." Unpublished Doctoral Dissertation. Tennessee State University, Nashville, TN. August.

Tomlinson, T.M. (1988). Class Size and Public Policy: Politics and Panaceas. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Whittington, E.H. (1985). "Effects of Small Class Size (1:15) on the Teaching/Learning Process in Grade One." Unpublished Doctoral Dissertation. Tennessee State University, Nashville, TN. August.

_____, H.P. Bain, C.M. Achilles. (1985). "Effects of Class Size on First-Grade Students." ERS Spectrum, 3, 4 (Fall), pp. 33-39.